

## 遠隔地コミュニケーションのための実時間人物仮想化

國田, 政志  
九州大学システム情報科学研究院知能システム学部門

羽山, 大介  
九州大学システム情報科学研究院知能システム学部門

吉松, 寿人  
九州大学システム情報科学研究院知能システム学部門

有田, 大作  
九州大学システム情報科学研究院知能システム学部門

他

<https://hdl.handle.net/2324/5924>

---

出版情報 : 火の国情報シンポジウム, 2004-03  
バージョン :  
権利関係 :

# 遠隔地インタラクションのための実時間人物仮想化

國田政志†、羽山大介†、吉松寿人†、有田大作‡、谷口倫一郎‡

†九州大学大学院システム情報科学府

‡九州大学大学院システム情報科学研究所

〒816-8580 春日市春日公園 6-1

E-mail: {mkunita, daizz, hisato, arita, rin}@limu.is.kyushu-u.ac.jp

本稿では実時間人物仮想化という概念を提案する。人物仮想化とは、人物に関する情報のうち必要なものを抽出しシンボルと呼ばれる情報で表現する事、さらに、その逆にシンボル群から人物の表現をつくり出すことを目的とする。複数の人物が仮想環境を介してインタラクションを図るためには、利用者に煩雑な操作を強要することなく必要な情報を積極的に取得するシステムが求められる。ここでは、まず仮想空間インタラクションを実現するための実時間人物仮想化の流れを説明する。次に、そのプロトタイプシステムとして現在我々が構築している仮想講義環境 VEIDL について述べる。このシステムは入力装置に実時間モーションキャプチャシステムを利用する。入力された体の動き情報をジェスチャ認識によってシンボル化し、コンピュータグラフィクスで描画される仮想環境中のアバタに、対応した動作を振舞わせる。最後にこのシステムの動作状況を示す。

**キーワード** 人物仮想化、仮想環境、遠隔地インタラクション、リアルタイムヒューマンプロキシ

## Real-time Human Virtualization for Distance Interaction

Masashi KUNITA, Daisuke HAYAMA, Hisato YOSHIMATSU,

Daisaku ARITA, Rin-ichiro TANIGUCHI

Department of Intelligent Systems, Kyushu University

6-1 Kasuga-koen, Kasuga, 816-8580 Japan

E-mail: {mkunita, daizz, hisato, arita, rin}@limu.is.kyushu-u.ac.jp

In this paper, we propose Real-time Human Proxy (RHP); a concept of real-time human-related information virtualization for avatar-based interaction. The objective of RHP is to obtain essential information, as form of symbols, from a large amount of data measured by some input devices. First, we describe RHP. Secondly, we show a prototype system of RHP. The system uses motion capture systems as input devices, symbolizes motion parameters by gesture recognition, and displays avatars behaving in a virtual space. Lastly, experimental results are shown.

**Key words** Human Virtualization, Virtual Environment, Distance Interaction, Real-time Human Proxy

## 1. はじめに

遠隔地を結ぶ実時間コミュニケーションの手段として、すでに電話が利用され始めてから長年の月日が経過している。現在ではコンピュータに関連した技術の進展により、実時間遠隔コミュニケーションの手段はさらに多様となっている。その例として挙げられるものは、キーボード入力による文字を送受信する「テキストチャット」、共通の描画領域を複数人で共有し、マウス入力による図を送受信する「ホワイトボード」、更にはカメラによって撮影した動画像を送受信する「ビデオチャット」など多岐に渡り、これらのコミュニケーションの手段は複合的に利用し得るものである。

これらの手段を用いた人と人とのインタラクションにおいて、電話やビデオチャットは最も豊かな表現力を持つ手段であると考えられる。これは人間にとって重要な視覚または聴覚に関わる情報を、ありのままに相手へと伝えられることが期待できるからである。その結果、これらの手段は対一のコミュニケーション手段として普及の一途を辿っている。一方、会議のように多人数が同時にコミュニケーションをはかるために用いる手段としては、電話およびビデオチャットは使いにくいものである。電話では、音声のみを頼りに複数人と会話を行なわなければならない。そのため、発話者を識別するのが困難であり、複数人が同時に発言すると聞き分けが難しくなる。また発話者が誰に注意を向けて発言しているかを知ることが困難であり、会話に参加している相手を全て把握することはほぼ不可能である。ビデオチャットでは、各チャット参加者はそれぞれ独立したウィンドウ内に表示されるため、複数人で使用する場合には、画面上に全ての参加者のウィンドウを並べて表示することになる。参加者が正面にカメラを配置するならば、画面に表示される参加者は全て正面を向いた姿で映るため、誰に注意を向けて発言しているかの判断は難しい。音声は電話と同様に全て同一チャンネルに合成さ

れるため、発話者を特定するには画面に映る参加者の中から発話者を捜し出さなければならない。

電話やビデオチャットが複数人のコミュニケーションの手段に適していないのは、人と人との空間情報が欠落していることに由来する。いずれの手段も一方の人物に纏わる視覚および聴覚情報を計測し、これを他方にそのまま提示することによってコミュニケーションを図ろうとするものである。この際、コミュニケーションを取る相手が単数である場合には、情報を受ける側がその存在位置どこかに仮定することで問題は起こらない。しかし、相手が複数に及ぶ場合、相手が自分に対してどの位置に存在するかという個々の空間的配置に関する情報が得られず、その整合性がとれないため、実際に一堂に会した場合には起こらない混乱を招くことになる。

コミュニケーションの参加者全てがそれぞれ別々の遠隔地に存在する場合、これを解決するためには空間的配置を仮想的に決定する手法が必要となる。その手法として、全ての参加者をアバタとして仮想環境に投影し、参加者は仮想環境中に表示されるアバタを介してコミュニケーションを図る研究が行なわれている<sup>[1][2]</sup>。しかしこの手法では、現在の参加者の状態を絶えず仮想環境に反映させなければならない。そのためこの手法には、全ての参加者にそれぞれのアバタを的確に操作し続けなければならないことを強要し、さらにその操作は、時に参加者に対して本来のコミュニケーションの目的とは関係のない負担を強いることになるという問題がある。

本稿では、仮想環境を用いたコミュニケーションを想定し、人物を仮想環境に投影するために、実時間人物仮想化(Real-time Human Proxy)について述べる。実時間人物仮想化とは、参加者に特別な操作を強いることなく、普段の自然な振舞を積極的に取得し仮想化するシステムの概念である。このシステムでは入力装置としてモーションキャプチャシステムを用いる。取得される体の動きをシンボル化することによって仮想化を行ない、

これを参加者にわかりやすい動作として仮想環境に表示する。本稿中では、さらに筆者らが製作中の、この概念を用いたプロトタイプシステムを紹介する。

## 2. 実時間人物仮想化の概念

### 2.1 人物仮想化

人物仮想化とは、人物の情報をシンボル化する、またはその逆を行なうという概念である。人物の情報とは、例えば手、足の動きや形状、顔の形や表情、声質、発せられた言葉、無意識にしがちな癖など、その人に関する情報一般を指す。シンボルとは、それらのうち1部分の情報についての記述である。すなわち、人物仮想化とは、ある人に纏わる情報をシンボルという単純な記述の集合へ変換すること、もしくはその逆に、シンボル群からその人の表現をつくり出すことである。

人物に関する情報は、その全てが常に必要とされているわけではない。例を挙げるならば、人と人がインタラクションを行うことが目的であれば、必要とされている情報には、例えば話している相手、話の内容、向いている方向、表情、手の仕草、会話に対する興味や度合、その人の癖などの項目が挙げられる。また、これがロボットを遠隔操作することが目的であれば、必要とされる操縦者の情報は、見ようとしている方向、移動の方向や速度、手の位置、物を掴もうとしているかどうか、手の掴む力、姿勢などが挙げられる。一方で、ロボットの遠隔操作を行なう際に操縦者の表情は必要とされないかもしれない、人と人のインタラクションにおいては手の掴む力は必要とされていないかもしれない。伝達すべき情報とそうでない情報は想定する状況によって決定される。人物仮想化は人物に纏わる情報を適当と考えられる細かさのシンボルに分解し、必要と思われるものを選択する。そして選択されたシンボルからその人物を表現する。選択から表現の間にはネットワークが存在することもあり、表現にはコンピュータグラフィックスの形で画面に表示されることもあれば、ロボットの動きの形で示さ

れることもある。

### 2.2 実時間人物仮想化

実時間人物仮想化(Real-time Human Proxy、以下RHP)とは、人物仮想化を実時間で行なうものである。ここでは利用目的として、仮想環境を用いたアバタによるインタラクションを想定している。

仮想環境を用いたインタラクションでは、全ての参加者は、仮想環境にアバタとして参加する。参加者は、そのアバタを本人の代理としてインタラクションを行なう。すなわち、仮想空間中に現れるのは基本的にアバタのみであり、そのためアバタは参加者本人の状態を十分に反映するものでなければならない。

仮想空間を用いたインタラクションにおいて、RHPが主張する人物仮想化とは、実世界の人物を観測し仮想空間中のアバタとして表現する過程に合致する。仮想空間中に投影されるアバタはインタラクションを代理で行なうものであり、その役割はインタラクションの質を左右する重要なものである。一方、的確に参加者の意図をアバタへ反映させることは一般に困難である。よって、RHPに求められる機能は以下ようになる。

- 参加者を実時間で仮想化する
- 参加者に積極的な入力作業を強要しない
- 参加者の自然な振舞いを利用し、特別な操作を覚えさせなくてよい
- インタラクションに必要な動作をシンボル化する
- 理解しやすい形で表現する

以下では、これらの機能を実現するRHPの構成について述べる。

### 2.3 RHPの構成

RHPは人物情報の入力、シンボル化、伝達、表現の4過程に分解される。ここでは、体の動きについての入力、シンボル化、表現について述べる。

### 2.3.1 入力: モーションキャプチャシステム

入力が困難な人間の情報に体の動きがある。これはアバタを介したコミュニケーションには欠かせないものであるが、人間のからだの動きが持つ自由度は非常に高いため、キーボードやマウスのような入力装置では十分な入力を確保することが難しい。加えて、体の動きを表現することがインタラクションの目的そのものでない場合も多く、この場合は参加者が煩雑な入力操作によってアバタの動作を決定しなければならないことは大きな負担となるばかりでなく、本来目的とするインタラクションの妨げともなり得る。

RHPでは体の動き情報に関して、実時間モーションキャプチャシステム(以下MCS)を入力装置として使用する。これにより体の動きそのものを処理することが可能となり、普段の仕草をそのまま入力として利用できる。したがって参加者は、入力方法を特別に学習する必要がなく、負担を大きく軽減できる。

### 2.3.2 シンボル化: ジェスチャ認識

MCSから得られる人間の動き情報は動作に関するシンボルへと変換される。変換にはジェスチャ認識の手法を利用する。認識されるべきシンボルは想定するインタラクションに合わせてあらかじめ決定しておく。これは想定するインタラクションの性質、および設定によって、必要とされる動作情報の種類が変わってくるためである。

### 2.3.3 表現: 動き情報生成

シンボルの形で表現された人物の情報を利用者に分かりやすく提示するために、コンピュータグラフィクスでアバタを表示する。あらかじめシンボルに対応づけた動き情報を用意しておき、それを基にアバタの動きを生成し、アバタに振舞わせる。より自然な動作を表現するために、動き情報はモーションキャプチャを利用して用意することも可能である。

## 2.4 RHPの利点

上に述べたRHPの構成によって、仮想環境インタラクションの利用者に負担をかけず

アバタを操作することが可能となる。MCSを利用することで、利用者は普段の自然な体の動きを用いてアバタを動かすことができるため、入力装置の特殊な操作法を覚えたり、煩雑な操作をインタラクションの間に強要されることがない。利用者が意識せずとも体の動きが入力されるため、キーボードやマウスのような装置を使用したときのように敢えて自分の動作を入力するという動作を必要とすることもない。これはアバタの動きを操作すること自体が目的ではないインタラクションにおいて、利用者の負担を大きく軽減できることになる。また、送るべきシンボルの種類が多い場合でも、ホタンやメニューを増やし、操作性を悪化させる必要はない。

ジェスチャ認識によってシンボル化を行うことで、インタラクションに必要な情報を抽出して通信を行うことができる。MCSから得られる姿勢情報全てを利用者間で送受信するのに比べて無駄がなく、利用者により分かりやすい形で提示することが可能となる。

シンボル情報からアバタの動きを生成することで、表示側の表現の自由度が高くなる。これは、アバタの動きを事前に用意した知識から生成するためである。MCSから得られる情報をそのまま用いてアバタを動かす場合はアバタの寸法は正確に利用者に合わせておかななくてはならない。さらに、MCSから得られないかもしれない細部の情報、例えば、表情や指の曲げ具合、爪先の角度や髪揺れなどは再現されず、結果不自然な描写に陥るか、その補完に労力が必要とされる。動作情報を事前知識から生成することで、アバタの寸法を合わせる必要もなく、細部に渡って動作を自然に生成することも容易になる。

## 3. RHPのプロトタイプシステム: VEIDL

我々はRHPのプロトタイプとして仮想講義環境(Virtual Environment for Immersive Distributed Learning、以下VEIDL)を作成している。VEIDLは講義参加者全員がそれぞれ遠隔地から接続する仮想環境(図1)であ

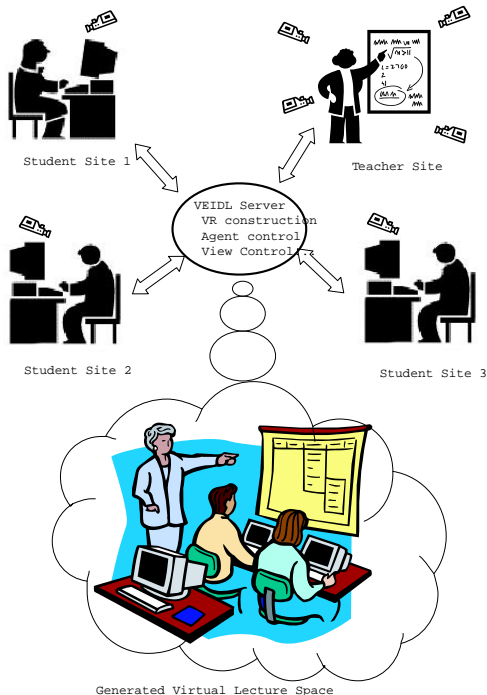


図 1: VEIDL の想定する仮想講義

る。全ての参加者は仮想空間に登場するアバタを通じて他の参加者とインタラクションを行う。

### 3.1 システム概要

VEIDLでは、各参加者はそれぞれ入力装置としてカメラとマイク、出力装置としてディスプレイとスピーカをもつ。カメラによって得られた参加者の非言語情報はシンボル化され、他の各参加者へ配信される。各参加者は、受信したシンボルに基づいて仮想環境をコンピュータグラフィクスによって描画し、それをディスプレイに表示する(図2)。仮想環境は参加者のアバタが存在する位置からの視点で描画される。マイクから得られる音声情報はそのまま各参加者に配信され、発言者アバタの仮想位置をもとに参加者のアバタの視点に合わせて立体音響を生成し参加者に提示する。これにより参加者は自分のアバタの視点から描画された他の参加者のアバタを通じてインタラクションを行うことが可能となる。

以下に非言語情報の入力から表示までの流

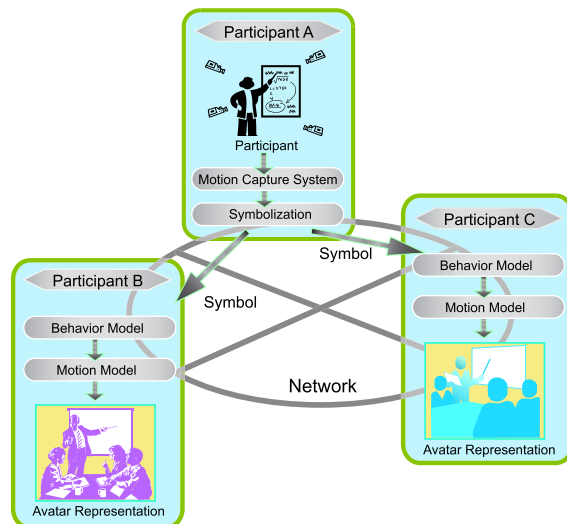


図 2: VEIDLにおける非言語情報の流れ

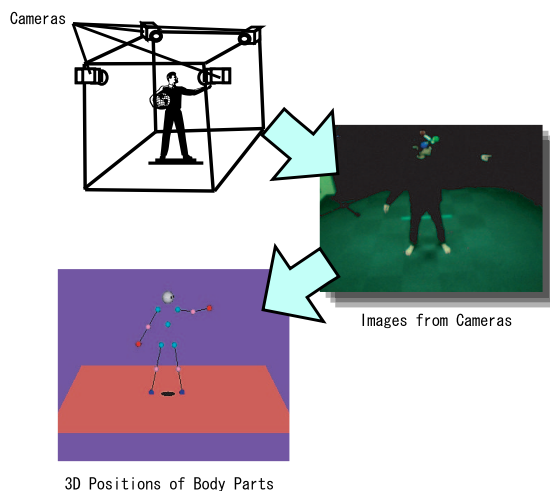


図 3: 全身用非接触型実時間MCS

れを述べる。

### 3.2 動作情報の獲得

情報の獲得には、現在我々が構築している実時間非接触型のMCSを用いる。使用するMCSには全身型(図3)と上半身型(図4)の2種類がある。全身型MCS<sup>[3]</sup>は10台のカメラと、PCクラスタ、および我々が構築した実時間並列画像処理環境RPV<sup>[4]</sup>によって成り立つ。このシステムによって、体にマーカーをつけることなく実時間で頭部、両手、両足、

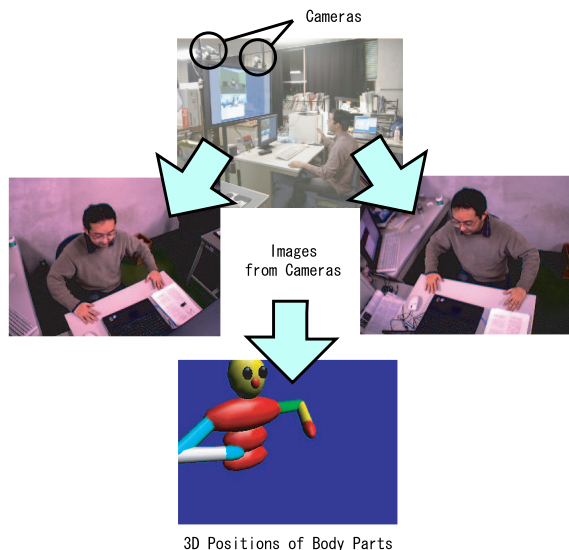


図 4: 上半身用非接触型実時間 MCS

両肘、両膝、および胴体の3次元座標を求めることができる。これは、講師側の情報獲得に利用する。一方、上半身型MCSは<sup>[5]</sup>対象人物の正面に配置された2台のカメラと1台のPCによって成り立つ。このシステムはデスクトップインタフェースを想定しており、体にマーカーをつけることなく実時間で頭部および両手の3次元座標を求めることができる。これは、生徒側の情報獲得に利用する。

得られた体部位の3次元座標をもとにジェスチャ認識を用いてシンボル化する<sup>[6]</sup>。VEIDLで使用するシンボルは、動きの種類を表すラベルと、それに関連する値とで構成される。講義環境で必要な動作情報として、特に必要であると考えられる以下の4種類のシンボルを採用する。それは、‘pointing with finger at’、‘looking at’、‘walking’、‘turn body to’である。これらのシンボルのうち方向を指定する値をとまなうものは、その値として対象となる参加者の識別子を用いる。MCSから得られた情報を基に、これらの動作が検出された場合にはそのシンボルを他の参加者に送信する。

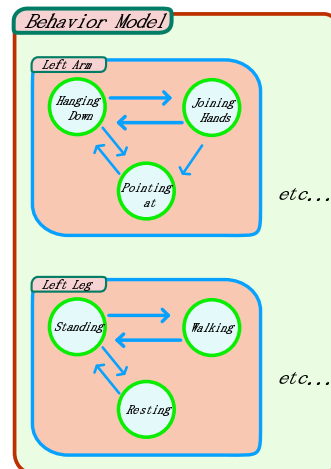


図 5: 行動モデル

### 3.3 通信

システムの実時間性を考慮して、通信には基本プロトコルとして Real-time Transport Protocol<sup>[8]</sup>を使用する。これは、実時間情報の送信を想定して作られたプロトコルであり、マルチキャストを用いることができる。またその制御プロトコルである RTCPは遠隔会議を想定したユーザー管理機能を備えている。これにより、各参加者のシンボルと音声情報は他の全ての参加者に配信される。

### 3.4 動き生成

最後に、送られてきたシンボルから、対応した動作を生成してアバタに振舞わせる。ここでは、動き情報を行動モデルと動きモデルの二つのモデルを用いて生成する。

#### 3.4.1 行動モデル

行動モデルとは、図5に示すような状態遷移図である。アバタの動作は行動モデルによって決定される。状態遷移図のそれぞれの要素は特定の動作を示すものであり、その動作とは、例えば手を挙げる、指で指す、歩く、腕を組む等の、意味を持った一連の動きの単位であり、一つのシンボルに対応づけられる。要素の間を結ぶ状態遷移確率は、動作の移り変わりやすさを示し、これにより不自然な動作のつながりを防ぐことができる。シンボル

が受信されると、状態遷移確率を操作しシンボルに対応した動作に状態遷移させることでその動作をアバタに行わせる。また、あらかじめ決められているシンボルのみでは、利用者がシンボルに対応づけられている動作をしない間は、アバタが完全に静止することになるため、これを回避するために、行動モデルは、腕を組む、ポケットに手をいれる等、直接インタラクションに影響しない動作を含んでいる。これらの動作はシンボルの送信されない間、状態遷移確率にしたがって発動される。

### 3.4.2 動きモデル

行動モデルによって決定されたアバタの動作は、動きモデルによって実際のアバタの動きへと変換される。動きモデルとは、特定の動作に対応した、動作を行う際に描く体の部位の軌跡やその速度情報の集合である。動きモデルは実際にアバタがどのように動くかを詳細に記述したものであり、動きを当てはめるアバタの人体モデルに密接な関係がある。そのため、アバタのモデルはあらかじめ用意しておき、そのモデルに対応した動きモデルを使用してアバタの動きを生成する。

## 3.5 表示

生成された各参加者の動きを用いて、仮想環境をコンピュータグラフィクスで描写する。この際、ディスプレイとしてHMD（ヘッドマウントディスプレイ、図6）を用いる場合、頭部の動きに合わせて視点、および視線を動かすことができる。

## 4. システムの動作状況

### 4.1 環境

RHPによって仮想化された人物を示すために、ここではVEIDLにおいて講師の動作がシンボル化され仮想環境のアバタに投影される様子を示す。講師アバタ1人と生徒アバタ4人によってVEIDLを動作させ、講師役の参加者が全身用MCSを使用して仮想環境に参加する。参加者は、シンボルとして登録されている動作と登録されていない動作をMCSの中で行い、その結果仮想環境に表現



図 6: HMD

されるアバタの様子を合わせて観察する。講師のRHPによる仮想化の様子を示すことが目的であり、他の参加者とのインタラクションは行わない。そのため、生徒側のアバタはただ仮想空間に配置するのみとして人物の仮想化は行わないものとする。図7にその結果を示す。

## 4.2 結果

図7中の各項目について以下に述べる。(a)は参加者がシンボルとして登録されていない動作を行った場面である。その動作はシンボル化されず、その結果アバタには動作は反映されていない。シンボルが送信されないため、振舞いモデルによって‘両手を組む’動作が発動されている。(b)および(c)は参加者が前方に歩行を開始した場面である。動作情報が‘walking’シンボルに変換され、仮想環境のアバタに反映されている。(d)および(e)は生徒を指さしている場面である。同様に‘pointing with finger at’シンボルがによってアバタに反映されている。(f)はシンボルとして登録されていない手の動作を行った場面である。手の動作はシンボル化されず、仮想環境のアバタには反映されていない。一方、顔の向きが生徒の方向であることが‘looking at’にシンボル化されており、アバタに反映されている。



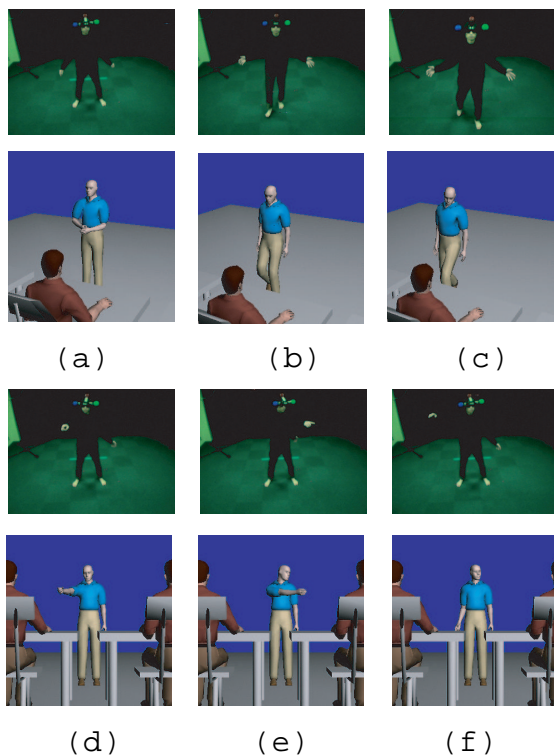


図 7: VEIDLにおける実時間人物仮想化の様子: MCSの中の参加者(上)および仮想環境での表示(下)

## 5. おわりに

本稿では、人物仮想化の概念を述べ、続いて仮想環境インタラクションを想定したRHPについて述べた。また、RHPのプロトタイプシステムであるVEIDLについて述べ、その動作状況を示した。その結果、実時間でVEIDLが人物を仮想化する様子が示された。

## 謝辞

本研究は科学研究費補助金学術創成研究(2)「人間同士の自然なコミュニケーションを支援する知能メディア技術(13GS0003)」の補助を受けた。

## 参考文献

- [1] Phillip Jeffrey, Andrew McGrath, “Sharing serendipity in the workplace”, Proceedings of the third international con-

ference on Collaborative virtual environments, pp.173-179, 2000.

- [2] M. Roussou, A. Johnson, T. Moher, J. Leigh, C. Vasilakis, and C. Barnes, “Learning and building together in a virtual world”, Presence, vol. 8, no. 3, pp. 247-263, Jun. 1999.
- [3] N. Date, H. Yoshimoto, D. Arita, S. Yonemoto and R. Taniguchi, “Performance Evaluation of Vision-based Real-time Motion Capture”, Proc. Workshop on Parallel and Distributed Computing in Image Processing, Video Processing, and Multimedia, in IPDPS CD-Rom Proceedings, 2003.
- [4] 有田大作, 花田武彦, 谷口倫一郎, “分散並列計算機による実時間ビジョン”, 情報処理学会論文誌 Vol.43 No.SIG 11 (CVIM 5), 2002.
- [5] Satoshi Yonemoto and Rin-ichiro Taniguchi, “A Direct Manipulation Interface with Vision-based Human Figure Control”, Proc. HCI International 2003, pp.811-815, 2003.
- [6] 吉松寿人, 有田大作, 谷口倫一郎, “没入型分散講義における講師についての非言語情報のシンボル化”, 電子情報通信学会総合大会 p.265, 2003.
- [7] 羽山大介, 花田武彦, 吉本廣雅, 有田大作, 谷口倫一郎, “没入型分散講義における自然な振る舞いをする講師アバタの実現”, 電子情報通信学会総合大会 p.264, 2003.
- [8] <http://www.cs.columbia.edu/hgs/rtp/>